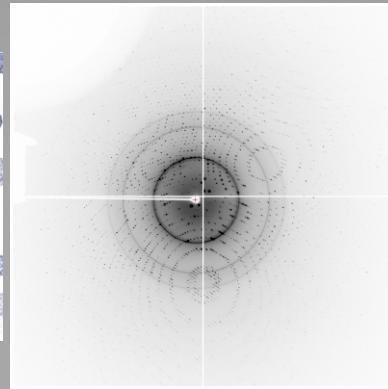
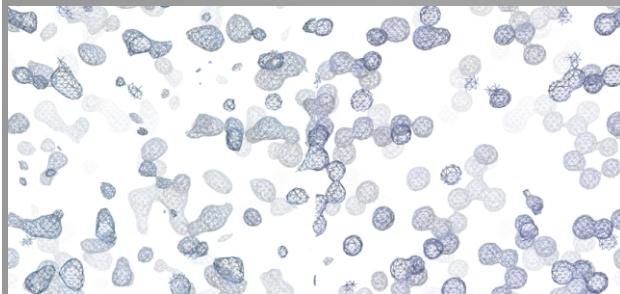
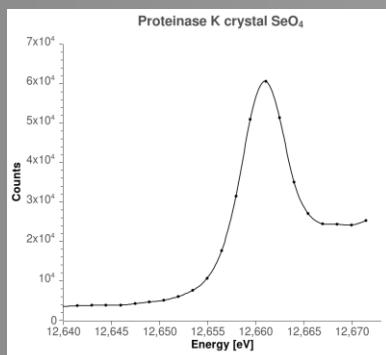
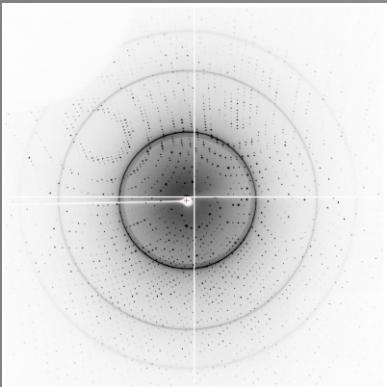


# Selenate in Macromolecular Crystallography



# PLAN

- Overview of standard phasing methods
- PDB survey
- Selenate and previous use
- Crystallization and phasing of 2 model proteins
- Comparison phasing results of sulfate and  $\text{SeO}_4$
- Soaking experiment in selenate
- Conclusions and future work

# Standard phasing methods

- Isomorphous replacement (SIR, MIR)
- Anomalous diffraction (SAD, MAD)
- Isomorphous replacement with anomalous scattering (SIRAS, MIRAS)
- They all require the presence of heavy atom.  
naturally present in metalloprotein (Cu, Co, Zn, Fe ...) or introduced (Se, Br, Hg, Pt, Ho, Au, Cd ...)
- Sulfur SAD phasing
- Radiation damage induced phasing (RIP, RIPAS)
- (Molecular replacement / direct methods)

# Phasing Methods : Pros / Cons

- S : native sample / low E = radiation damage, low  $\delta f$ "
- Br<sup>1</sup><sub>13.4 keV</sub> : quick, simple / not systematic
- Classic soaking in HA (Pt, Hg, Au, U ...) : high  $\delta f$ "  
1  $\lambda$  / lot of testing, samples, data (specific interaction, damages)
- SeMet<sub>12.6 keV</sub> : efficient / requires immobile Met, time
- 1990<sup>2</sup> : 1<sup>st</sup> protein crystal structure determined by SeMet reported | the most used phasing method

<sup>1</sup>Dauter *et al.*, JMB, 1999 ; <sup>2</sup>Hendrickson *et al.*, EMBO J.

# PDB survey 02/2011

- Total / Crystal structures : 71138 / 61856

Precipitant /  
transport/ binding

<b>K<sup>+</sup></b>	<b>1032</b>	3.6 keV
<b>Na<sup>+</sup></b>	<b>2931</b>	1.0 keV
<b>Cl<sup>-</sup></b>	<b>4567</b>	2.8 keV
<b>Ca<sup>2+</sup></b>	<b>5365</b>	4.0 keV
<b>Mg<sup>2+</sup></b>	<b>5835</b>	1.3 keV

Metalloprotein /  
purification

<b>Br<sup>-</sup></b>	<b>193</b>	13.5 keV
<b>Co<sup>2,3+</sup></b>	<b>391</b>	7.7 keV
<b>Ni<sup>2,3+</sup></b>	<b>519</b>	8.3 keV
<b>Cu<sup>1,2,3+</sup></b>	<b>823</b>	8.9 keV
<b>Fe<sup>2,3+</sup></b>	<b>1734</b>	7.1 keV
<b>Zn<sup>2+</sup></b>	<b>6361</b>	9.6 keV

Precipitant /  
cryo

<b>TAR</b>	<b>57</b>
<b>CIT</b>	<b>408</b>
<b>MPD</b>	<b>603</b>
<b>EDO</b>	<b>2047</b>
<b>GOL</b>	<b>5269</b>

Precipitant

<b>PO<sub>4</sub><sup>3-</sup></b>	<b>2440</b>	2.1 keV
<b>SO<sub>4</sub><sup>2-</sup></b>	<b>8331</b>	2.4 keV

Most common energy range in standard MX beamlines  
 $11.0 \pm 3.5$  keV (7.5-14.5 keV) (1.9-0.85 Å)

Elucidate function / phasing

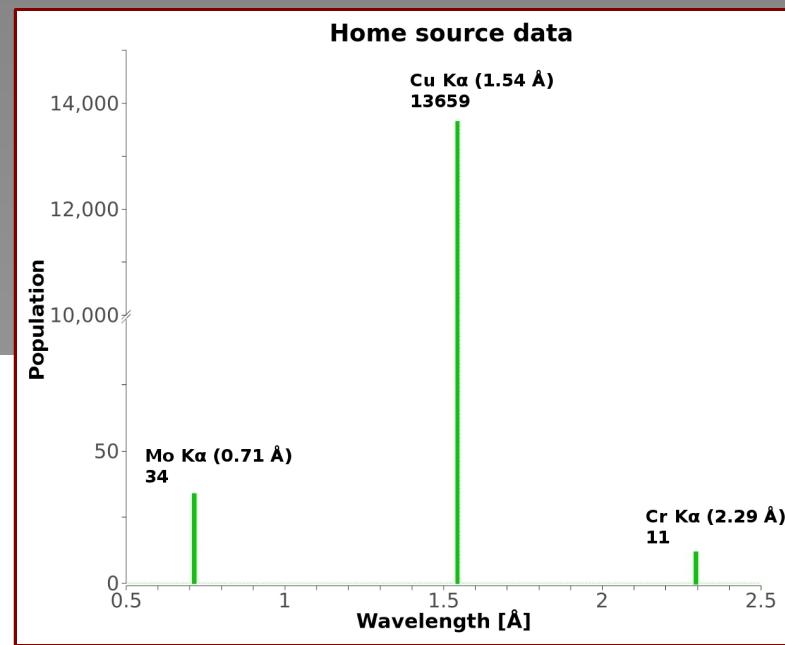
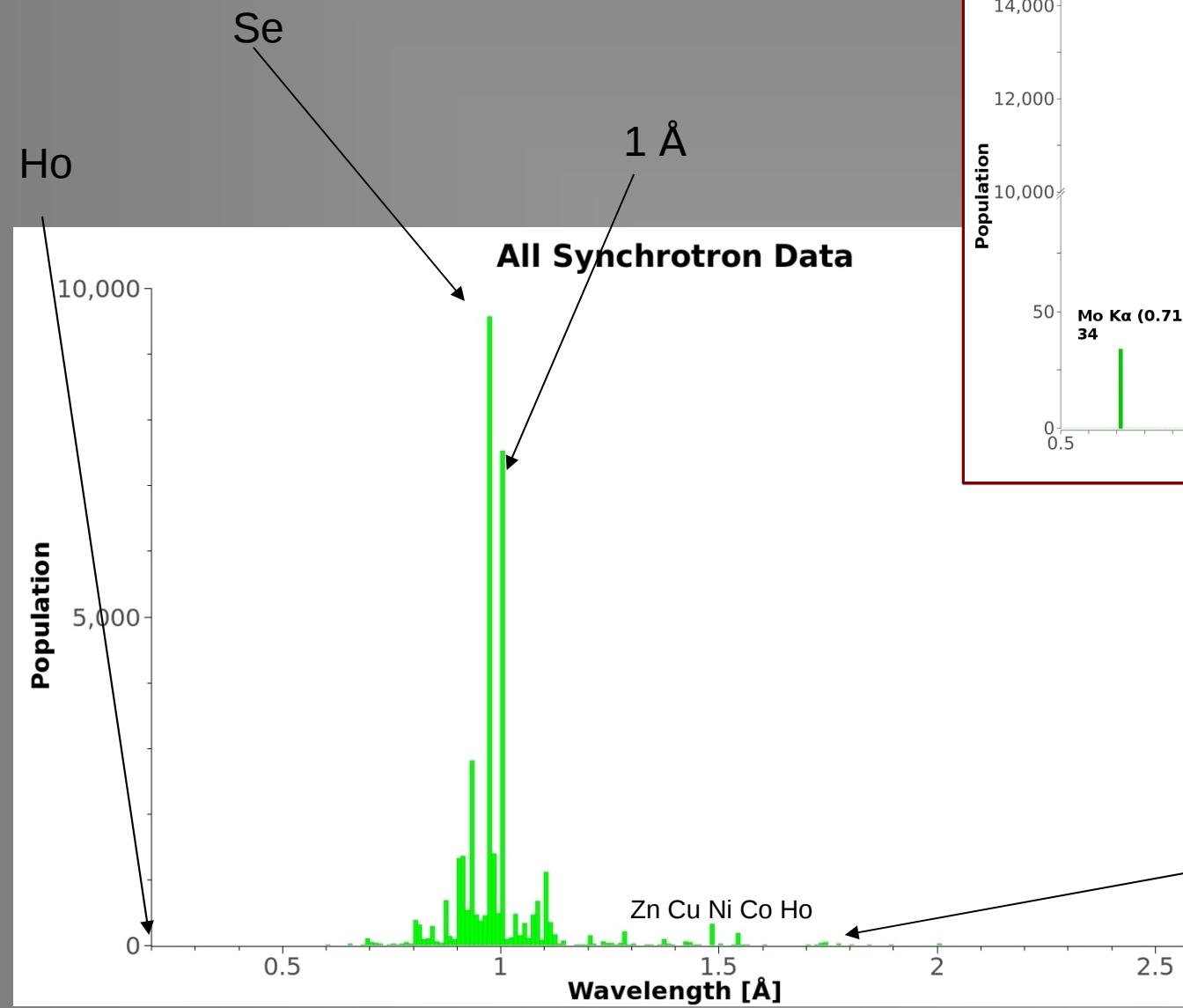
<b>Rb<sup>+</sup></b>	<b>31</b>	15.2 keV	(Na <sup>+</sup> )
<b>Ba<sup>2+</sup></b>	<b>56</b>	5.2 keV (LIII)	(Ca <sup>2+</sup> )
<b>Sr<sup>2+</sup></b>	<b>88</b>	16.1 keV	(Ca <sup>2+</sup> )
Ca <sup>2+</sup> /Mg <sup>2+</sup> : Ho <sup>3+</sup>	7	Yb <sup>2,3+</sup>	55 Tb <sup>3,4+</sup>

P : As Sb ?  
 Lower number / Toxic

-> S : Se ?

lithium 3 <b>Li</b> 6.941	beryllium 4 <b>Be</b> 9.0122	boron 5 <b>B</b> 10.811	carbon 6 <b>C</b> 12.011	nitrogen 7 <b>N</b> 14.007	oxygen 8 <b>O</b> 15.999	fluorine 9 <b>F</b> 18.998
sodium 11 <b>Na</b> 22.990	magnesium 12 <b>Mg</b> 24.305	aluminum 13 <b>Al</b> 26.982	silicon 14 <b>Si</b> 28.086	phosphorus 15 <b>P</b> 30.974	sulfur 16 <b>S</b> 32.065	chlorine 17 <b>Cl</b> 35.453
potassium 19 <b>K</b> 39.098	calcium 20 <b>Ca</b> 40.078	gallium 31 <b>Ga</b> 69.723	germanium 32 <b>Ge</b> 72.61	arsenic 33 <b>As</b> 74.922	seelenium 34 <b>Se</b> 78.96	bromine 35 <b>Br</b> 79.904
rubidium 37 <b>Rb</b> 85.468	strontium 38 <b>Sr</b> 87.62	tin 49 <b>In</b> 114.82	antimony 50 <b>Sn</b> 118.71	tellurium 51 <b>Sb</b> 121.76	iodine 52 <b>Te</b> 127.60	lutetium 53 <b>I</b> 126.90
caesium 55 <b>Cs</b> 132.91	barium 56 <b>Ba</b> 137.33	radium 87 <b>Fr</b> [223]	radium 88 <b>Ra</b> [226]			

# PDB distribution 12/2009



# PDB structures with $\text{SO}_4$

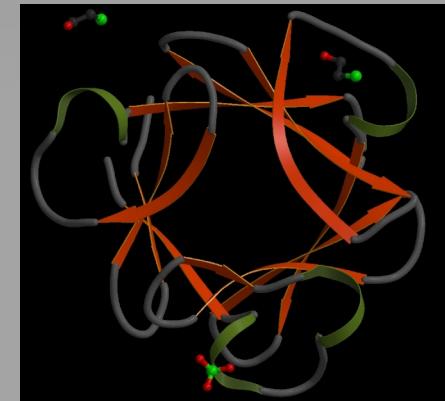
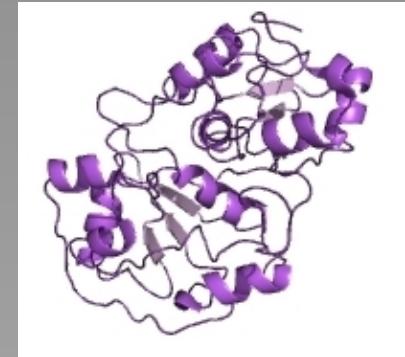
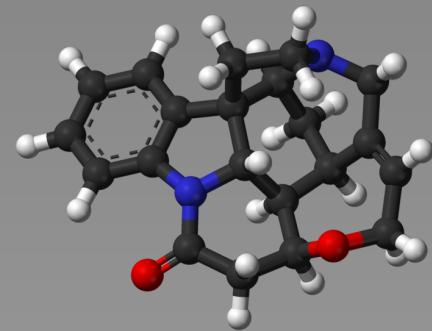
- 8331 structures : > 13 % of all crystal structures
- From 0.78 Å to 4 Å
- As small as 12 AA with 1  $\text{SO}_4$  and 36 AA with 4  $\text{SO}_4$
- As large as 7280 AA (14 x 520) with 43  $\text{SO}_4$  and also with 22  $\text{SO}_4$
- Average MW = 30 kDa (2200 non-H atoms)
- Average number of residues per  $\text{SO}_4$  "site" = 240

# $\text{SeO}_4$

- Sulfate :  $\text{SO}_4^{2-}$  / Selenate :  $\text{SeO}_4^{2-}$
- Structurally /chemically similar
- S-O bonds : 1.5 Å, O-S-O angles : 108-111 °
- Se-O bonds : 1.6 Å, O-Se-O angles : 107-113 °
- $\text{SeO}_4^{2-}$  diameter is 0.2 Å larger compare to  $\text{SO}_4^{2-}$
- S K-edge : 2472 eV | Se K-edge : 12658 eV
- $(\text{NH}_4)_2\text{SO}_4$ ,  $\text{Na}_2\text{SO}_4$ , Ca., Cu., ...
- $(\text{NH}_4)_2\text{SeO}_4$ ,  $\text{Na}_2\text{SeO}_4$ , Ba., ..

# Past Impact of Selenate in Crystallography

- 1950<sup>1</sup> : structure determination of Strychnine a small molecule of 330 gr.mol<sup>-1</sup> (*SIR*)
- 1984<sup>2</sup> : confirm 2 SO<sub>4</sub><sup>2-</sup> at binding site in Bovine Liver Rhodanese (32 kDa) (*MIRAS*)
- 1993<sup>3</sup> : confirm the presence of SO<sub>4</sub><sup>2-</sup> at the active site of Human Basic Fibroblast Growth Factor (17 kDa) (*MR difference map*)



<sup>1</sup>Bokhoven *et al.*, Acta Cryst. ; <sup>2</sup>Lijk *et al.*, Eur. J. Biochem. ; <sup>3</sup>Eriksson *et al.*, Protein Sci.

# Motivations

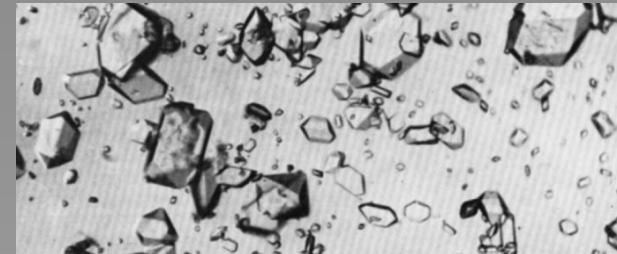
- Study crystallization of proteins in  $\text{SeO}_4$
- $\text{SO}_4^{2-}$  is the most used anion in precipitants
- > 13 % of crystal structures contain  $\geq 1 \text{ SO}_4$
- $\text{SeO}_4$  can be dissolved to  $\sim 3.5 \text{ M}$  in  $\text{H}_2\text{O}$  @ RT
- In theory a rather simple substitution
- Most of fixed E beamlines above the Se K-edge
- Most of future beamlines  $\sim \text{Se}$

# Proposed study

- At least 2 standard samples for crystallization study
- Definition of “Standard” protein sample:  
low cost and easily “crystalizable”  
more specifically: in  $\text{SO}_4$  and preferentially with  $\text{SO}_4$  bound
- Most common protein sample for these studies:  
Hen Egg White Lysozyme : HEWL  
But : no HEWL structure with  $\text{SO}_4$  bound

# Proteinase K

- E.C.3.4.21.64 ; endopeptidase K; 28.9 kDa
- 279 AA; 5 C (2 S-S); 5 M, 2018 atoms
- Purif./Cryst. reported in 1974<sup>1</sup>
- Structure solved in 1984<sup>2</sup> (MIR)
- 25<sup>(+27unrel.)</sup> PDBs : 0.83 to 2.9 Å
- Mostly @ 2.2 Å
- Crystallization condition: mainly  $\text{CaCl}_2$  /  $\text{NaNO}_3$
- 1  $\text{SO}_4$  bound (2PWA) VDHD RT,  $\text{CaCl}_2$  &  $\text{NaNO}_3$ ,alanine boronic acid pH 6.5
- ProtK crystals grow in many conditions (+ evaporation)
- $\text{P}4_32_12$  UC ~ 68 68 107 Å; (SC 42 %)

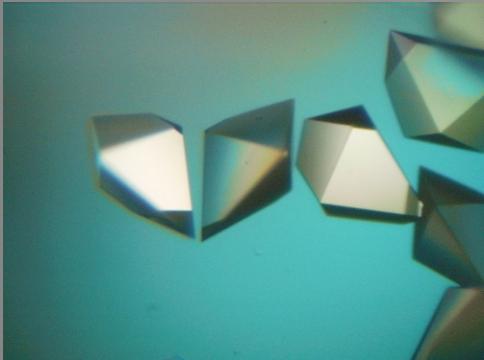


Ebeling et al., Eur. J. Biochem. (1974)  
... crystallized readily even from crude solutions if the ionic strength was low and the protein adjusted to 40-50 mg/ml, forming bipyramidal prisms ...

<sup>1</sup>Ebeling et al., Eur. J. Biochem. ; <sup>2</sup>Pahler et al., EMBO J.

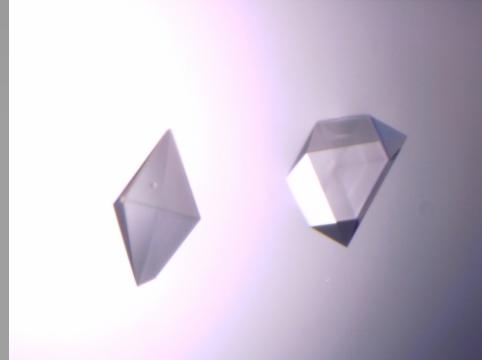
# Proteinase K crystallization

- Vapour diffusion hanging drop method at room temperature
- From 5 to 500 mM  $\text{Na}_2\text{SeO}_4$  (all conditions gave crystals)
- ProtK\* @ 40 mg/ml in  $\text{H}_2\text{O}$  ; Vwell = 900  $\mu\text{L}$  and Vdrop = 1 + 1  $\mu\text{L}$
- Additional screen performed with  $\text{Na}_2\text{SO}_4$  for site comparison



50 mM  $\text{SO}_4$  (5 days)

200  $\mu\text{m}$



50 mM  $\text{SeO}_4$  (2 days)

Cryo-solution consisted of mother liquor + 25 % glycerol

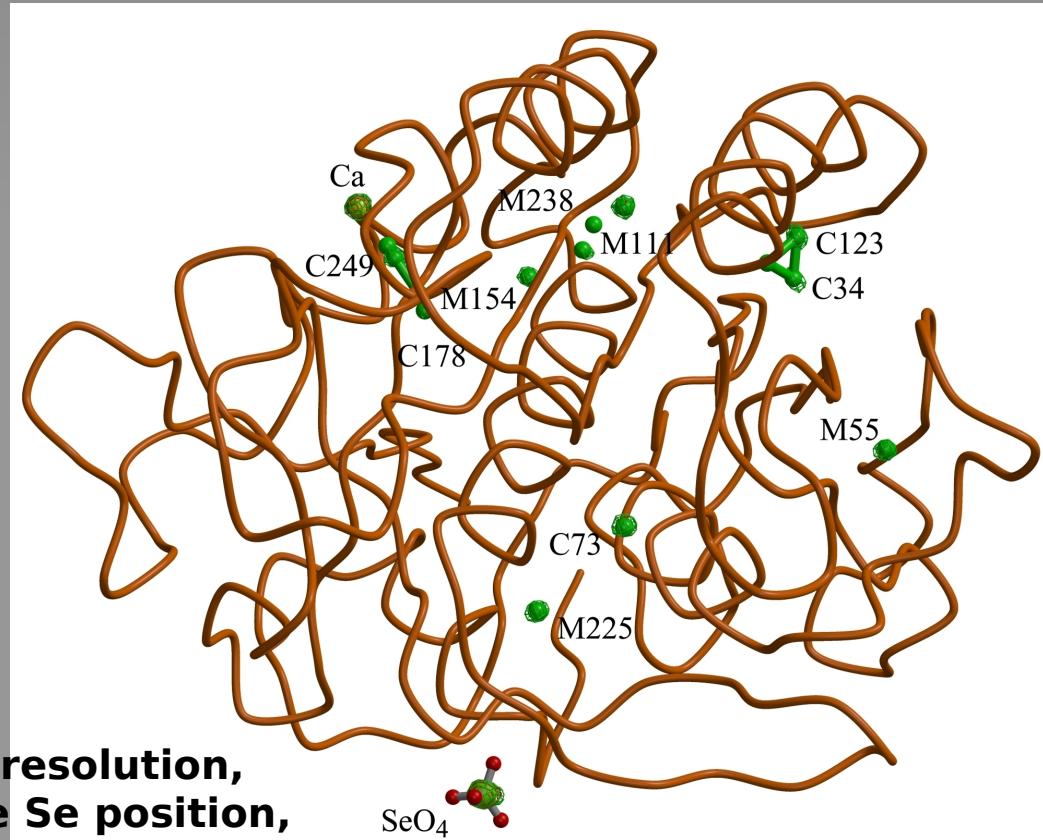
# Proteinase K data collection

- Data collected at NSLS beamline X6A , E = 14.5 keV (0.8551 Å)

<b>Space group</b>	P4 <sub>3</sub> 2 <sub>1</sub> 2
<b>Unit Cell (Å) (a b c)</b>	68.05 68.05 102.55
<b>Resolution Limits (Å)</b>	20.00-0.94 (0.95-0.94)
<b>Completeness (%)</b>	98.7 (85.6)
<b>I / σI / R merge (%)</b>	30.5 (2.0) / 4.6 (29.8)
<b>Average mosaicity (°)</b>	0.19
<b>Multiplicity (180 °)</b>	12.1(4.0)
<b>Wilson B factor (Å<sup>2</sup>)</b>	4.3

Anomalous difference map  
(3.5 σ in green and 7.5 σ in red)

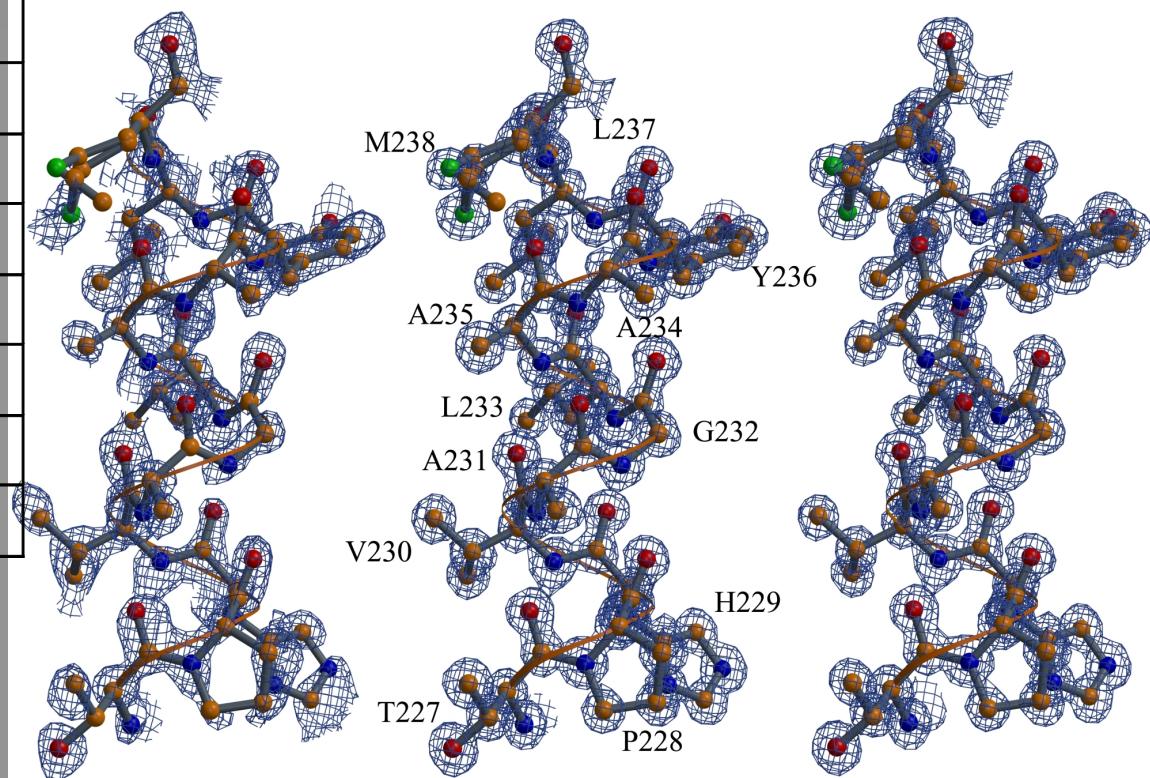
**Map calculated with the 0.94 Å resolution,  
1<sup>st</sup> peak (94 σ) is located at the Se position,  
2<sup>nd</sup> peak ( 36 σ) the peak height at the Ca site  
All Cys S and Met S with peak > 8.5 σ (including  
2<sup>nd</sup> conformation of Met 238 with 5.6**



# Proteinase K model building / refinement

Resol. D/ E (Å)/ cycl. DM	1.7 / 0.95 / 30
Contrast <sup>#</sup>	0.50 (0.32)
Connectivity <sup>#</sup>	0.83 (0.82)
Pseudo-free CC <sup>#</sup> (%)	90.3 (52.9)
CC MAP (%) (SAD/ SX- DM)	48.1 / 88.1
Residues properly placed	279
Resolution (Å)	20.0-0.94 (0.97-0.94)
R/R <sub>free</sub> (%)	9.0 / 10.3 (17.8/ 19.4)
B (prot, wat, lig) (Å <sup>2</sup> )	5.7 / 14.7 / 8.8
Favor./ Additional/ Other (%)	88.5 / 11.5 / 0.0
DPI (Å) / (PDB ID)	0.013 / 2V8B

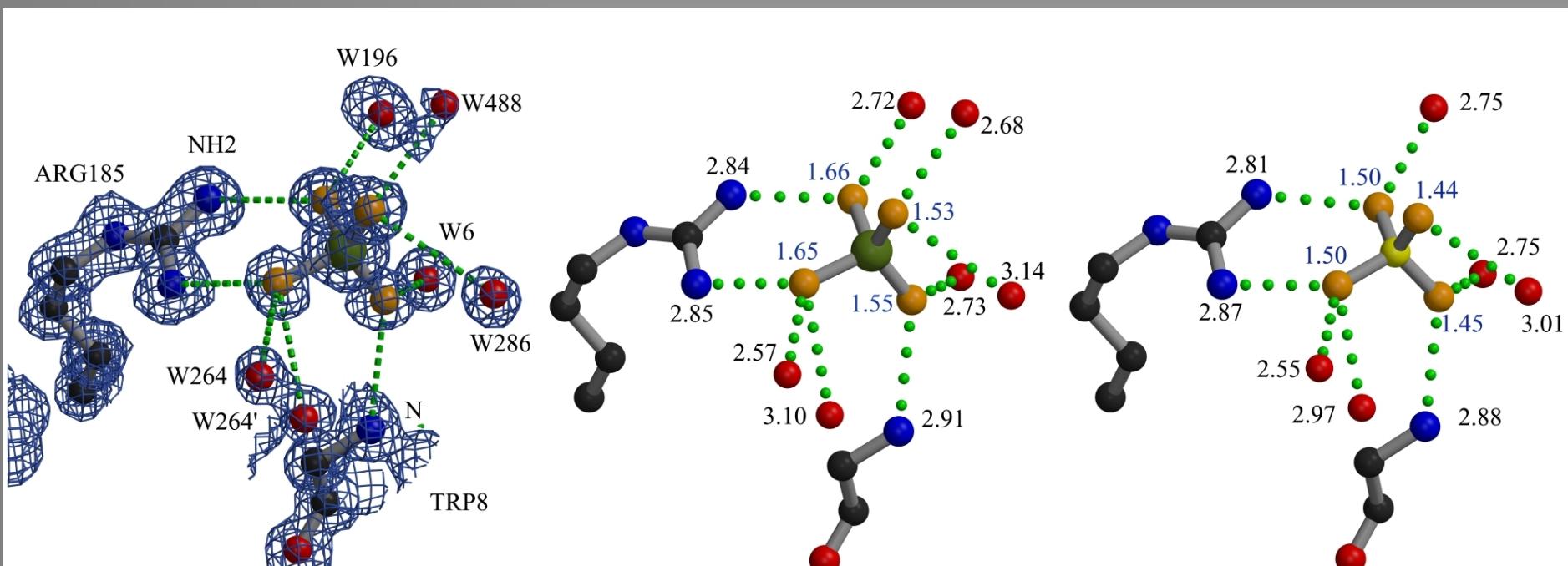
Refined model :  
279 AA, 1 SeO<sub>4</sub>, 1Ca, 497 W



# Proteinase K $\text{SeO}_4$ site

The selenate (occupancy of 0.6) lies at the surface of the protein, between ARG185 and TRP8.

ProtK  $\text{SO}_4$  structure  
refined at 0.95 Å  
data collected at 1 Å



2 Fo-Fc at + 1.5 $\sigma$   
Fo-Fc at + 4 $\sigma$   
Fo-Fc at -4 $\sigma$

SeO<sub>4</sub> bond length  
Hydrogen bond  
length

SO<sub>4</sub> bond length  
Hydrogen bond  
length

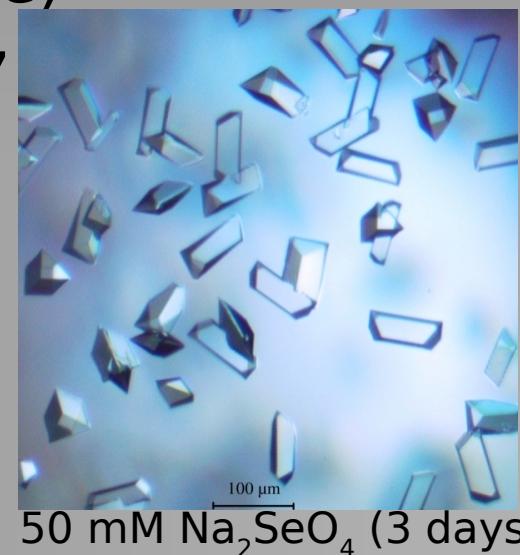
# Porcine pancreatic elastase

- E.C.3.4.21.36; serine proteinase; 25.9 kDa
- 240 AA; 8 Cys (4 S-S); 2 Met; 1822 non-H atoms
- Crystallization reported in 1968<sup>1</sup>
- 3.5 Å Structure solved in 1970<sup>2</sup> (MIR)
- 89 PDBs : 0.94 to 3.3 Å, mostly @ 2 Å
- Crystallization condition : NaAcetate / Na<sub>2</sub>SO<sub>4</sub>
- 71 PDBs with SO<sub>4</sub> bound, few with 2 SO<sub>4</sub>
- P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> UC ~ 51.5 58.0 75.5 Å (SC 40 %)

<sup>1</sup>Shotton *et al.*, JMB ; <sup>2</sup>Watson *et al.*, Nature

# PPE crystallization

- “Standard” condition : VDHD, RT, 20 mg/ml PPE, 10 mM NaAcetate pH 5.0, 20 mM Na<sub>2</sub>SO<sub>4</sub>
- From 5 to 200 mM Na<sub>2</sub>SeO<sub>4</sub> in 20 % ethylene glycol
- PPE\* @ 40 mg/ml (H<sub>2</sub>O), Vwell = 900 µL, Vdrop = 1 + 1 µL
- Best crystals : 50 mM (number and size)
- Seeding also performed using “native” µ-crystals grown in 40 mM SO<sub>4</sub>



# PPE data collection

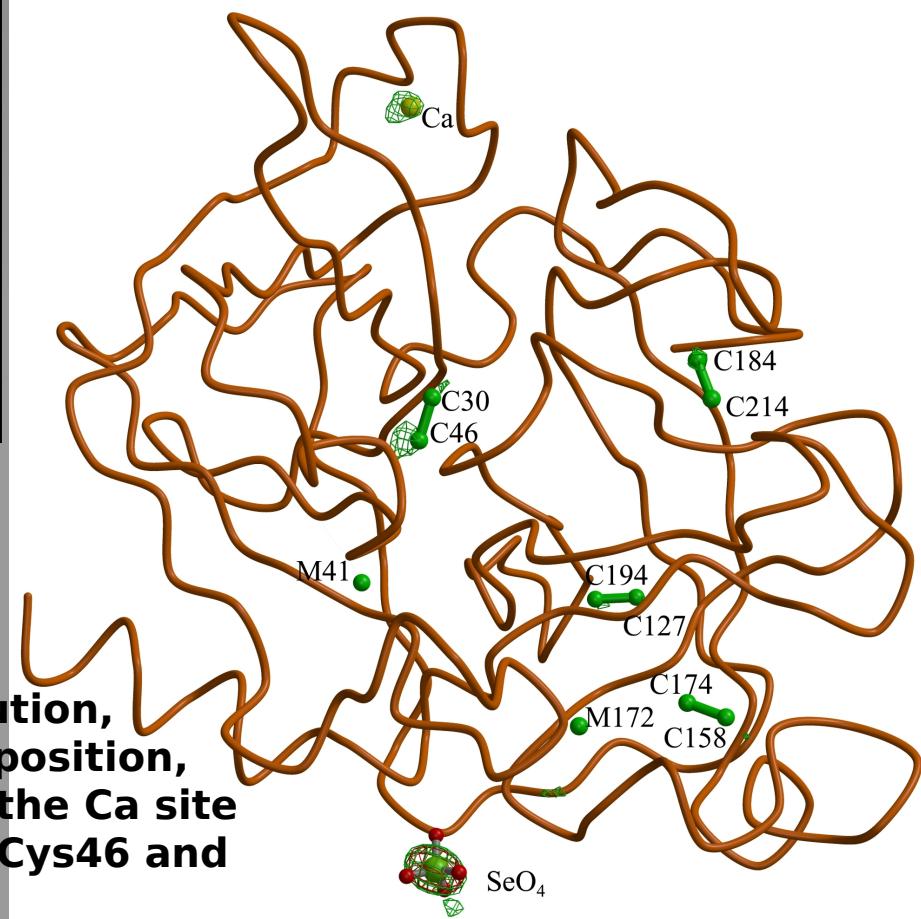
- Data collected at NSLS beamline X6A , E = 12.67 keV (0.9786 Å)

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit Cell (Å) (a b c)	514157.47 74.18
Resolution Limits (Å)	30.00-160 (166-160)
Completeness (%)	96.7 (77.4)
I / σI / R merge (%)	18.5 (2.1) / 5.3 (216)
Average mosaicity (°)	0.35
Multiplicity (156 °)	5.3 (2.3)
Wilson B factor (Å <sup>2</sup> )	14.5

Anomalous difference map  
(3.5 σ in green and 7.5 σ in red)

**Map calculated with the 1.6 Å resolution,  
1<sup>st</sup> peak (90 σ) is located at the Se position,  
2<sup>nd</sup> peak (5.5 σ) the peak height at the Ca site  
3<sup>rd</sup> 4<sup>th</sup> peaks (3.8 and 3.7 σ) at S of Cys46 and**

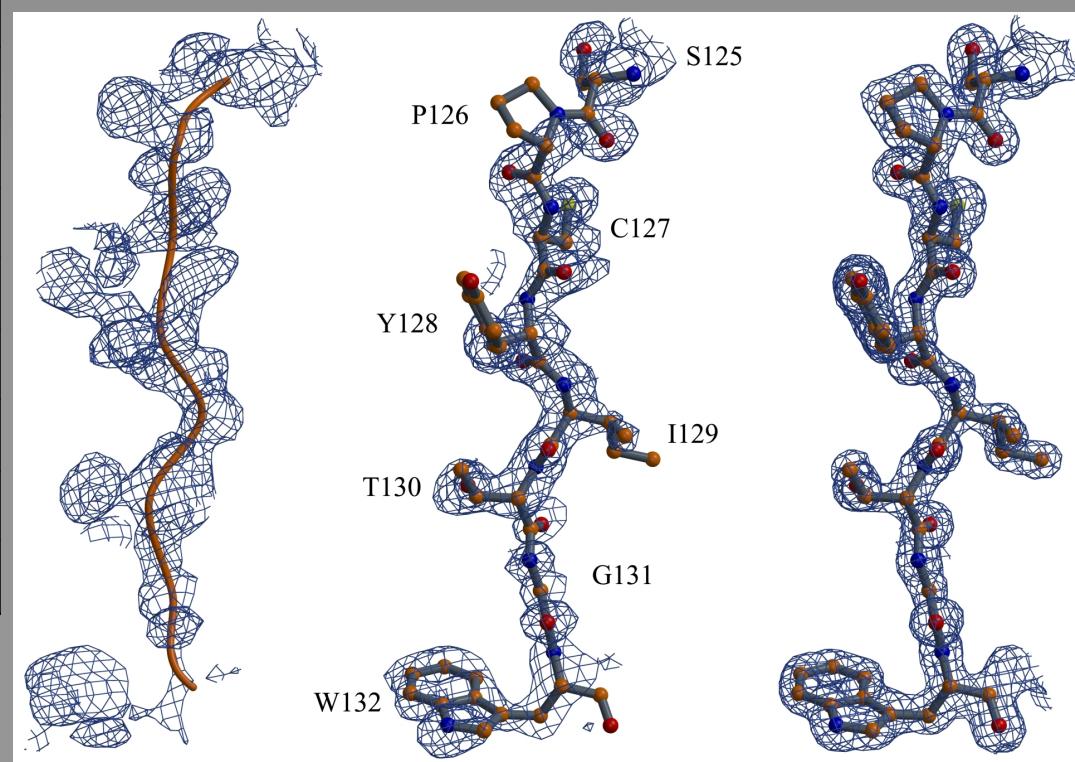
**Cys184**



# PPE phasing / refinement

<b>Resol. D/ E (Å) / cycl. DM</b>	2.0 / 165 / 30
<b>Contrast*</b>	0.45 (0.35)
<b>Connectivity*</b>	0.89 (0.86)
<b>Pseudo-free CC* (%)</b>	69.5 (59.3)
<b>CC MAP (%) (SAD/ SX- DM)</b>	49.5 / 68.7
<b>Residues properly placed</b>	239
<b>Resolution (Å)</b>	30.0-165 (171-165)
<b>R/Rfree (%)</b>	17.5 / 20.6 (27.1/ 29.8)
<b>B (prot, wat, lig) (Å<sup>2</sup>)</b>	15.3 / 28.5 / 20
<b>Favor./ Additional/ Other (%)</b>	85.9 / 14.1/ 0.0
<b>DPI (Å) / PDB ID</b>	0.089 / 2V0B

Refined model :  
 240 AA, 1 SeO<sub>4</sub>, 1Ca, 277 W



SHELXD (no DM)  
 maps contoured at 1.5 σ

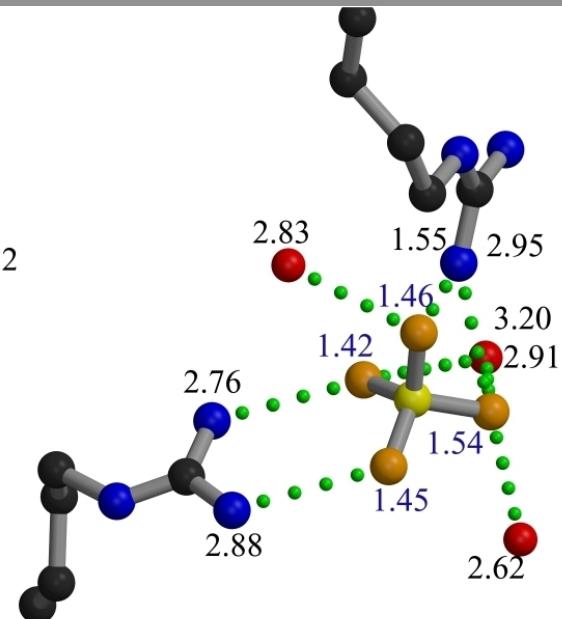
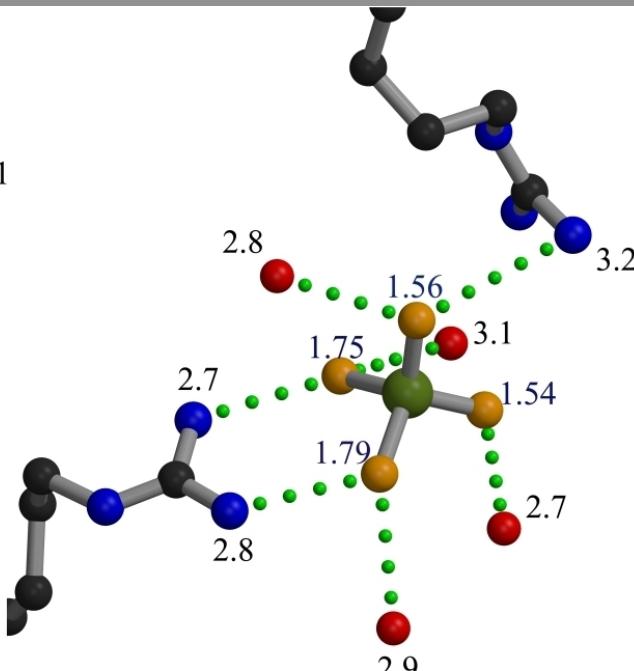
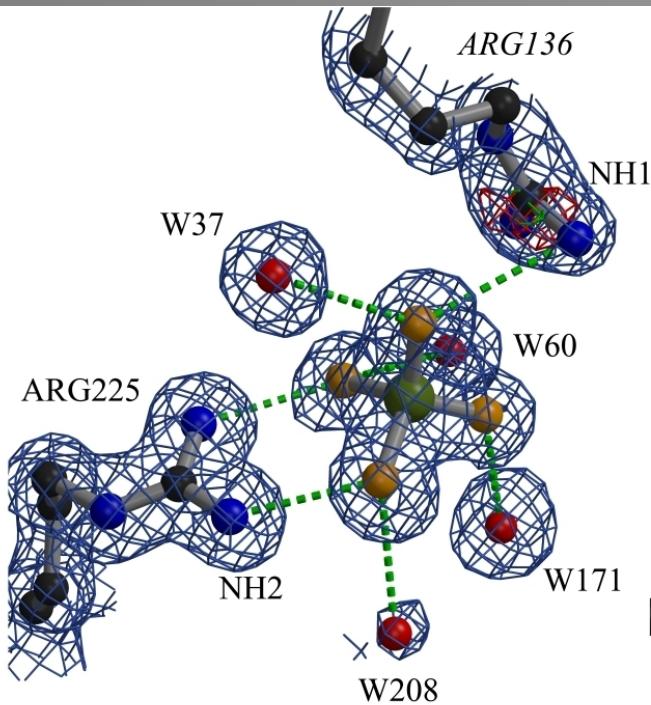
DM

REF 2Fo-Fc

# PPE $\text{SeO}_4$ site

The selenate (occupancy of 1) lies at the surface of the protein, between ARG225 and ARG136 of a symmetry related molecule.

PPE  $\text{SO}_4$  structure used :  
1GVK<sup>1</sup> (0.94 Å)



2 Fo-Fc at + 1.5 $\sigma$   
Fo-Fc at + 4 $\sigma$   
Fo-Fc at -4 $\sigma$

$\text{SeO}_4$  bond length  
Hydrogen bond length

$\text{SO}_4$  bond length  
Hydrogen bond length  
<sup>1</sup>Katona et al., JBC (2002)

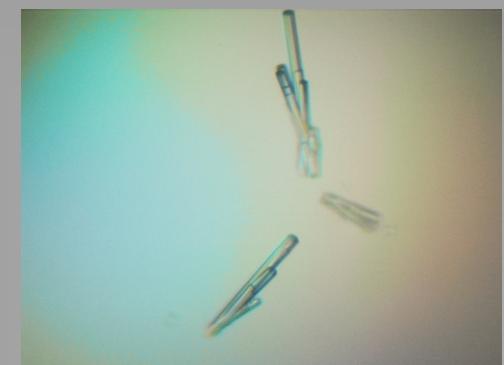
# $\text{SeO}_4$ soaking experiment ?

- X6A user protein (210 AA) : 1.2 M Am.Sulfate, 0.2 M NaCl, pH 8 (Tris-HCl), 20 % Glycerol
- Native crystals diffract to 1.95 Å ( $\text{P}3_1\text{2}_1$ )
- Soaked 12 H in 1.2 M  $\text{Na}_2\text{SeO}_4$  pH 8, 0.2 M NaCl, 20 % Glycerol
- $\text{SeO}_4$  sample diffracted to 2.1 Å
- Anomalous map indicated 1 Se site with poor occupancy at the known  $\text{SO}_4$  site (from final MR solution)
- $\text{SeO}_4$  did not displace  $\text{SO}_4$ ; Crystallization in  $\text{SeO}_4$  planned ?



$\text{SO}_4$

200 μm



$\text{SeO}_4$

# Co-Crystallization in SeO<sub>4</sub>

X6A user protein (360 AA / monomer, dimer in A.U.)

Condition : 0.6 M Am.Sulfate, pH 7 (Tris-HCl)

Native crystals diffract to 1.6 Å (P<sub>2</sub><sub>1</sub>2<sub>1</sub>2)

Protein did crystallized in Na<sub>2</sub>SeO<sub>4</sub> at pH 7

SeO<sub>4</sub> sample diffracted to 1.7 Å

Single data set above the Se K edge.

Anomalous map indicates 2 Se (1 site per monomer)

Confirmed with later MR solution, SAD map not conclusive.

Structure solved using SeMet and MAD phases.

**> 1 Se per mono. (low occup.) / 360 AA (B.R. <1 %)**

**MAD experiment at X6A planned / or sample with higher Se occ.**

# $\text{SeO}_4$ soaking experiment ?

Sample used, proteinase K.

Protein crystallized in a  $\text{SO}_4$ -less solution ( $\text{NaNO}_3$  published conditions).

Data collected at 6.85 keV at X6A | structure solved by S-SAD method.

Refined and the calculated anomalous difference maps showed no evidence of  $\text{SO}_4$ .

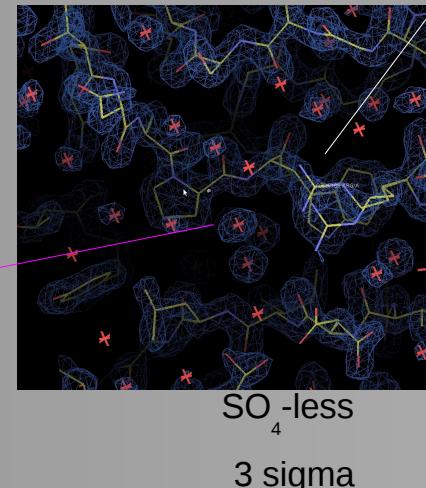
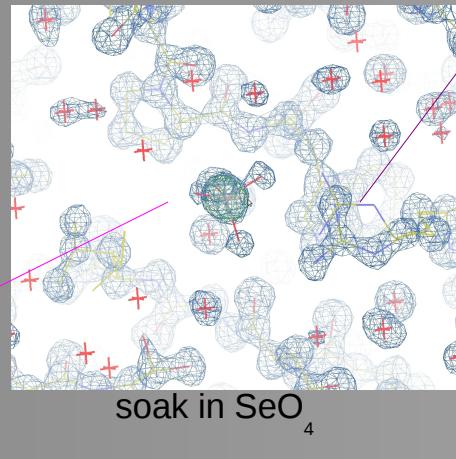
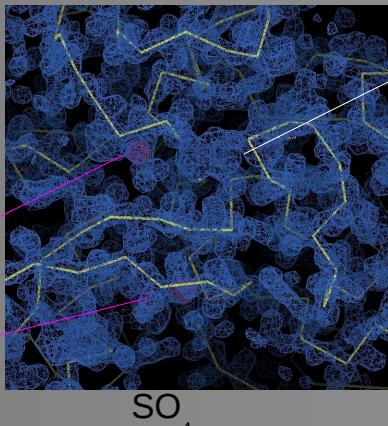
Protein crystallized in  $\text{SO}_4$ . Data collected below the Fe K edge (7 keV).

Structure solved by the S-SAD method (+ Ca). Maps clearly indicate presence of 1  $\text{SO}_4$ .

One crystal from the  $\text{SO}_4$ -less condition soaked in a  $\text{SeO}_4$  (0.33 M)

Data collected above the Se K edge.

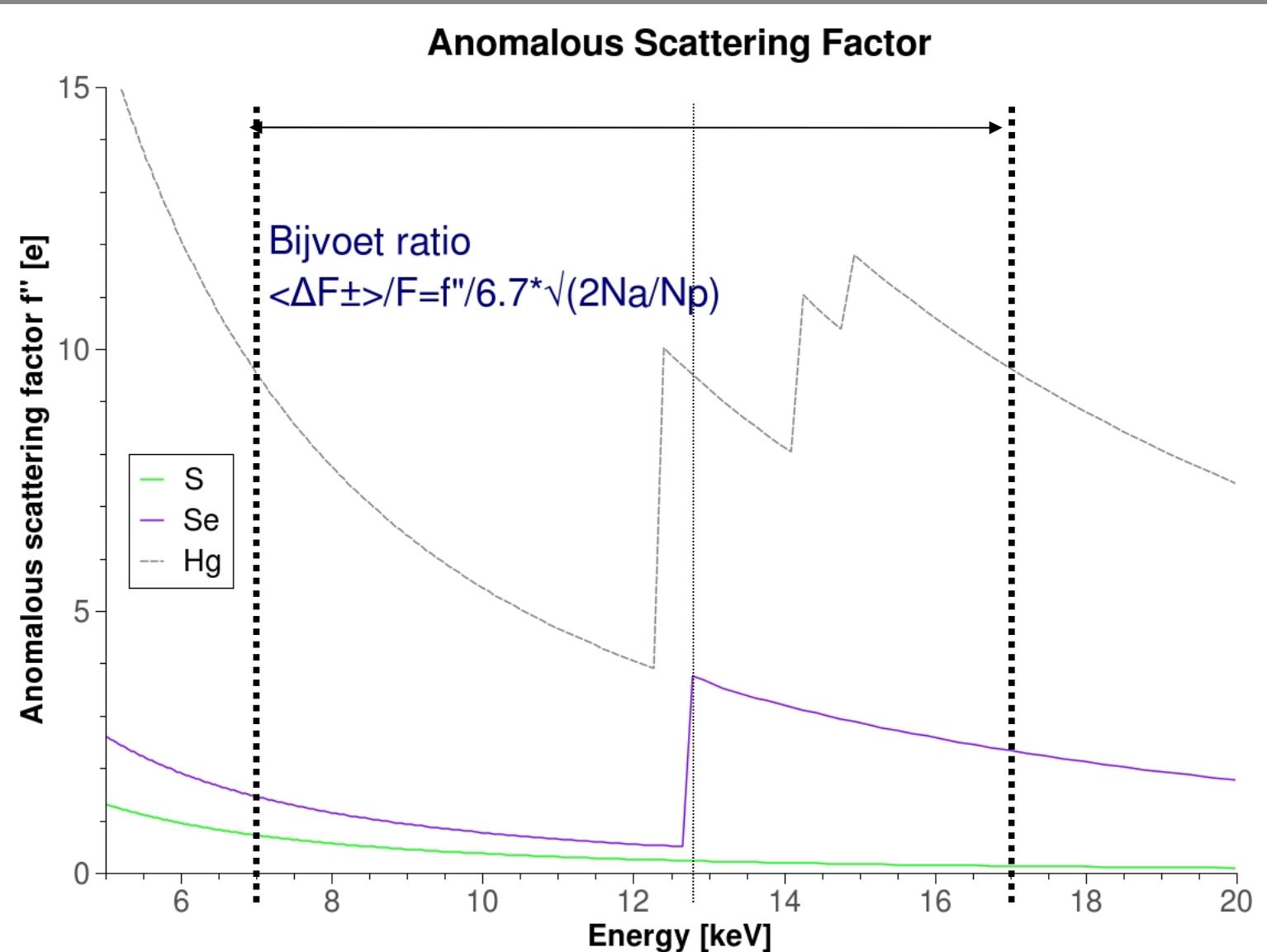
Maps indicate one clear and strong  $\text{SeO}_4$  ion, at the known  $\text{SO}_4/\text{SeO}_4$  position.



Same region  
ARG 185

# $\text{SO}_4 / \text{SeO}_4$

Bijvoet ratio<sup>1</sup>



<sup>1</sup>Hendrickson & Teeter, Nature (London)

# $\text{SO}_4$ / $\text{SeO}_4$

Estimated Bijvoet ratio and dose

Protein	Proteinase K		PPE	
Type	$\text{SO}_4$	$\text{SeO}_4$	$\text{SO}_4$	$\text{SeO}_4$
5 keV 2.48 Å	2.4	2.6	2.5	2.8
7.8 keV 1.59 Å	1.2	1.3	1.2	1.3
	$5 \ 10^4$	$5 \ 10^4$	$5 \ 10^4$	$5 \ 10^4$
12.7 keV 0.98 Å	0.4	1.9	0.5	2.0
	$2.3 \ 10^4$	$2.7 \ 10^4$	$2.4 \ 10^4$	$2.8 \ 10^4$

Doses based on experiments performed on one test sample and measured flux. Dose estimated for a cubic sample with size of 100x100x100  $\mu\text{m}^3$  and a beam size of 200x200  $\mu\text{m}^2$ .

Advantages : lowering the dose ( $\frac{1}{2}$ ) while increasing  $\Delta F$  ( $\sim 2$ )  
As  $\Delta F$  increase significantly, no necessity to record redundant data

# $\text{SO}_4$ / $\text{SeO}_4$

- Solve these structures relying on 11 S and 1 Ca ?

Attempts to solve the native structure of **PPE** with data collected at 12.7 keV failed.  
Resolution/multiplicity of data were : 1.5 Å / 7 (200 degrees recorded).

Peak heights (in  $\sigma$ ) in the anomalous map calculated with refined phases:

$\text{SO}_4$ : 8.2 ; Ca: 5.2 ; Cys194: 5 (for the Se data : 90, 5.5 and 3.8 (CYS184)).

→ increase multiplicity / dose ; lower energy to increase  $\Delta F$  ; increase phasing power

## Data collected on a “native” ProtK crystal

Energy (keV) / Oscillation range (°)	7.8 / 180	12.7 / 140
Anomalous signal (%) (PHENIX)	2.1	2.6
Resolution (Å)	30.0-17	15.0-0.95
Multiplicity / Rmerge (%)	11 / 2.7	5 / 3.8
CONTRAST / CONNECT / CC	0.34 / 0.88 / 0.65	0.42 / 0.81 / 0.52
Eantiomorph (SHELXE)	0.32 / 0.85 / 0.54	0.41 / 0.80 / 0.51
Residues built / R / Rfree(ARP)	269 / 19.8 / 24.0	NA*

Low energy data required to calculate phases.

High energy / resolution for refinement.

Increase multiplicity at high energy; increase phasing power.

30 trials in SHELXD, 50 cycles of DM in SHELXE.

\*Failed due to missing data at low resolution and low redundancy

# Conclusions

- Crystallization in  $\text{SeO}_4$  for 2 model proteins
- Structures solved by SAD: 1 Se for 280 AA
- Crystallization preferred over quick soaking (competition ..)
- Seeding using “native” crystals proven to be successful
- Very simple, .... not to try if  $\text{SO}_4$  precipitant (competition)
- Pretty inexpensive and quick
  
- $\text{SeO}_4$  occupies same sites as  $\text{SO}_4$ : native data ?
- $\text{SeO}_4$  soaking of ProtK sample grown in SO4-less : OK
- Quick soak in very high concentration of  $\text{SeO}_4$  ?
- Thio-selenate ( $\text{Se}_2\text{O}_3^{2-}$ ) to increase phasing power ?



# Acknowledgments



Vivian Stojanoff

NSLS and NIGMS beamline X6A STAFF

